

REMARKS

Claims 43-68 and 70 are pending in the present application. In the Final Office Action mailed June 5, 2006, the Examiner rejected Claims 43-68 and 70 under 35 U.S.C. §103(a). Each rejection will be addressed in the order as listed below.

I. Claims 43-52, 54-57, 59-68 and 70 rejected under 35 U.S.C. §103(a) as allegedly being unpatentable over WO00/72114 (hereinafter, Carden) in view of U.S. Patent No. 6,349,296 (hereinafter, Broder)

II. Claims 53 and 58 rejected under 35 U.S.C. §103(a) as allegedly being unpatentable over Carden in view of U.S. Patent No. 6,029,167 (hereinafter, Evans) and further in view of Gehringer et al, 2000, ASEE/IEEE Frontiers in Education Conference, Session F1B, pp. 2-7 (hereinafter, Gehringer)

I. Claims 43-52, 54-57, 59-68 and 70 are not obvious under 35 U.S.C. §103(a) over Carden in view of Broder

The Examiner has made the following points regarding the lack of teaching in Carden:

Carden fails to expressly disclose:

- Wherein said originality checking application includes rules for:
 - Obtaining, **fingerprinting**, and storing on a database relevant documents from a variety of sources which might be copied
 - **Fingerprinting uploaded** papers to be checked for originality
 - **Comparing a fingerprint of a paper to fingerprints of relevant documents to identify possible matches**
 - Comparing said paper's **full text to a full text** of all said possible matches
 - Generating an originality report which highlights those portions of the paper which match portions of said relevant documents identified as possible matches; wherein said originality report further comprises a report of the level of duplication between said paper's full text and said full text of said possible matches

(Final Office Action, page 4).

However, the Examiner alleges that the Broder patent teaches what the Carden patent fails to disclose:

“Broder teaches the process of digital fingerprinting of entire documents by breaking the document into a series of tokens called Shingles. A shingle is a contiguous set of tokens that specifically identify the document and a relation to other documents to form a set of fingerprints. The fingerprints represent a sketch of a document a good estimate regarding the resemblance can be used to form a basic level to a very-high level of resemblance can be used while filtering the documents (Broder column 4, lines 17-25 and 65-67 and column 5, lines 1-5 and 34-55 and column 11, lines 55-65). Broder also teaches the output of a result by showing the ordered pair list where all of the documents with the shared shingle will be presented and were the level of resemblance is tracked and displayed in the track. Broder also teaches the storing or storage of over 100 million documents in the search engine, which is a database of searchable content (Broder column 12, lines 1-5) that can accept a new document for storage and allow the resemblance checking to occur.”

(Final Office Action, page 4)

The Examiner concludes:

“It would have been obvious to one of ordinary skill in the art, having the teachings of Carden and Broder before him at the time of the invention was made, to modify the system of Carden to incorporate the resemblance checking and result output as taught by Broder, in order to obtain a system that is able to incorporate a plagiarism checking mechanism along with an author submissions of new or existing documents. One would have been motivated to make such a combination because of the expressed teaching of Broder to use the technique by authors to detect copies of their work or even to detect minor changes or even in licensing terms (see column 12, lines 15-3) and Carden is a system used by authors to submit bodies of work to their peers for review.”

(Final Office Action, page 4)

The Examiner further suggests:

“Broder expressly teaches a process for comparing identical documents where the entire document is compared against previously fingerprinted documents and if an identical match has been made then the clustering system can eliminate one of the identical documents from search engine database and maintain a reference to where the deleted document originated. The Examiner interprets the teaching of Broder as explicitly teaching a process of comparing entire text of an original document to a possible match (see column 10, lines 22-30) where Broder states “For identical documents, fingerprint the entire document” and “for equivalent documents, fingerprint either the canonical form or the set of shingles sorted”. While the document may be eliminated from the database,

nonetheless a full document scan is performed when a new document is encountered.”

(Final Office Action, page 16)

The Applicants respectfully disagree with the Examiner’s rejections and interpretations of Broder.

i. The combination of Carden and Broder fails to teach “comparing said paper’s full text to a full text of all said possible matches”

A required element of claim 43 is, for example, an originality checking application including rules for comparing a paper’s full text with the full text of a paper identified as a possible match. As noted above, the Examiner admits that the Carden patent does not teach or describe an originality checking application as required in claim 43.

Despite the Examiner’s characterization, the Broder patent does not teach an originality checking application including rules for comparing a paper’s full text with the full text of a paper identified as a possible match, as found in claim 43. As defined by Broder, the patent only describes methods for comparing a representative selection (e.g., fingerprint or sketch) of a Web page with a portion of another Web page identified as a potential match. Broder’s method of comparing identical Web pages is to break them down into fingerprints, or digital tags, so full text comparisons are not performed. The Examiner states just that when saying “Broder teaches the process of digital fingerprinting of entire documents....” wherein the digital fingerprints are compared, not the full text of a paper.

It should be pointed out that Broder’s definition of a “document” can be found in column 3, lines 41-47 when describing the key aspects of his system (Figure 1). Broder has chosen to be his own lexicographer as allowed under MPEP 2111.01 (III) and defines a “document” to be; “The Web allows for the interchange of multimedia data records frequently in the form of documents, that are commonly called Web pages. Associated with each Web page is an address called a Universal Resource Locator (URL). The Web pages are maintained by the servers. Users of the clients typically access the documents using Web browsers such as the Netscape Navigator™, or the Microsoft Internet

Explorer™.” Broder further marks his “document” as being number 100 and carries this numbering system throughout the specification as referenced in the text and the Figures. As defined by Broder then, “documents” are only “Web pages”, that are well known in the art as formed using a computer hypertext markup language such as HTML or XML, thereby creating documents that are viewable using a Web browser, which matches his definition. Therefore, the “document” of Broder as he has defined it does not include word documents or those documents found in a Portable Document Format (PDF) that might be accessible through a linked File Transfer Protocol (FTP) associated with a particular Web page. As such, Broder teaches only the resemblance comparison of Web pages, and not papers or text submissions as found in the present application. Therefore, the resemblance comparisons of Broder are done on Web pages, and the resemblance comparison is just that—how much does a particular Web page “resemble” another Web page by comparing representative selections of each.

If one were to interpret Broder’s “documents” more broadly by ignoring the definition of Broder (which is impermissible as per MPEP 2111.01 since Broder has specifically defined the meaning of “document”), then Broder still does not teach the comparison of full text to full text papers. Broder teaches the comparison of fingerprints of documents, and not as the Examiner interprets that the “entire document is compared against previously fingerprinted documents...” (Final Office Action, page 16). The fingerprinting of a document is defined in Broder as; “Fingerprints are short tags for larger data objects.” (column 6, lines 58-59). Therefore, the “entire document” as characterized by the Examiner must first be itself fingerprinted, i.e. characterized using a short tag, prior to it being compared with previously fingerprinted documents. As such, the full text documents are not being compared against each other, as only a portion of any document is being compared one to the other. In fact, Broder describes the procedure for preparing a document for comparison:

“The set of shingles associated with a document is too large to manipulate conveniently. Hence the shingles [contiguous sequence of k tokens] are first reduced to unique identifications and common shingles are ignored. The unique identification can be computed as digital fingerprints. Then, a representative selection is made from the set of fingerprints. This selection constitutes a sketch of the document. The

sketches must be such that they lead to a good estimate of the resemblance of documents.”
(column 4, lines 18-25, emphasis added)

This is not a full text to full text comparison as claimed in the present application, but a comparison of selected bits, or representative selections, of that particular document. A digital fingerprint is not a verbatim rendering of a document, but is a digital tag given to that document that is unique to that document. Broder does not perform the further comparison as found in claim 43, wherein Broder does not compare those possible matches identified by the fingerprint comparison by further “comparing said paper’s full text to a full text of all said possible matches of said relevant documents identified as possible matches.” Indeed, the purpose of the Broder system is to avoid such a comparison.

The purpose of the Broder patent is to find Web pages that share some level of similarity so that they can be grouped together. Searches for exact matches are specifically excluded as being contrary to the method and an undesired time sink. The whole purpose and structure of the Broder patent’s technology is to look for some minimum amount of similarity (e.g., resemblance) so as to find related Web pages, but to explicitly avoid comprehensive comparisons (such as full text or image comparisons) so as to permit the review of millions of Web pages in a reasonable amount of time. Full text review of Web pages would sabotage the function of the Broder patent in terms of timeliness. As such, the Broder patent does not teach an originality checking application including rules for comparing a paper’s full text with the full text of a paper identified as a possible match, as required in claim 43. Combining the method of the Broder patent with the Carden patent thus does not provide an originality algorithm that carries out full text searches.

Neither the Carden nor Broder patents, alone or in combination, teach an originality checking application including rules for comparing a paper’s full text with the full text of a paper identified as a possible match, as found in claim 43. Thus, the cited references do not provide all of the elements of the presently claimed invention. The Examiner has not pointed to any teaching in either reference that provides this element. The Applicants request these rejections be withdrawn.

ii. The combination of Carden and Broder fails to teach “generating an originality report...wherein said originality report further comprises a report of the level of duplication between said paper’s full text and said full text of said possible matches”

A required element of claim 43 is the generation of an originality report “...wherein said originality report comprises a report of the level of duplication between said paper’s full text and said full text of said possible matches.” As noted above, the Examiner admits that the Carden patent does not teach or describe an originality checking application as required in claim 43. As discussed above, the Broder patent determines resemblance between Web pages through analysis of representative selections; not an analysis of the full text for each Web page. Likewise, the Broder patent technique looks for the existence of, but not the degree of, similarity. As the Broder patent does not determine resemblance between Web pages with the full text of each Web page, the Broder patent does not and cannot teach the generation of an originality report on the level of duplication between a paper’s full text and the full text of possible matches. The fact that the Broder patent’s algorithms eliminate frequently occurring shingles and eliminate entirely close matching documents, the Broder patent method, when combined with the Carden patent, is contrary to methods where the level of duplication is to be determined and reported.

The Examiner suggests that:

“Broder expressly teaches a real-time resemblance process where Broder teaches a process (column 11, lines 49-51) of performing a full document scan and sorting the number of shingles that match and based on the number of shingles, present the result as to the level of resemblance and combined with the teachings (column 10, lines 20-30) that full documents are scanned then the Examiner interprets Broder as scanning an entire document and looking up the shingles that match and then sorting the matches and presenting the results to the user, which is a form of an originality report.”

(Final Office Action, page 17)

The Applicants disagree with the Examiner’s characterization of Broder. As described in Broder, shingles are not full text papers, but representations of a Web page as defined by Broder to be a “document”. Therefore, Broder does not teach full text paper scans or comparisons of full text papers. Broder describes “full sketches” of each Web page

(column 11, lines 59-51), which is not the same as full text scans. Also, in column 10, lines 20-30, Broder once again does not suggest full text scans, but states that clusters of Web pages can be represented by a single fingerprint (i.e. represented by a unique tag). There is no comparison in Broder where the full text of a paper is compared to the full text of another paper. All Broder is teaching is the comparison of Web pages that are first reduced to representative selections of the Web page, such as shingles, fingerprints, or sketches.

iii. There is No Motivation to Combine the Broder Patent with the Carden Patent

One would not be motivated to combine the Broder patent with the Carden patent in the manner suggested by the Examiner. The cited references do not acknowledge or make reference to the need or desire of persons in the art to use a system for reviewing papers with an originality checking application involving a full text comparison between papers and generating an originality report based upon the full text comparison between papers. Likewise, the cited references do not allude to a need or desire of persons in the art to generate a system with an originality checking application configured to conduct a full text comparison between papers and generate an originality report based on the full text comparison. Therefore, a person of ordinary skill in the art, in view of the cited references, would not have been motivated to combine the Broder patent with the Carden patent in the manner suggested by the Examiner to arrive at the present invention. This is particularly true considering the fact that the Broder patent's techniques are contrary to the purpose and function of the claimed invention, as described above.

To support the rejection, the Examiner has asserted that "One would have been motivated to make such a combination because of the expressed teaching of Broder to use the technique by authors to detect copies of their work or even to detect minor changes or even in licensing terms (see column 12, lines 15-3) and Carden is a system used by authors to submit bodies of work to their peers for review." Office Action, page 3. The Applicants first note that the Examiner is simplifying the claimed invention. Claim 43 requires a full text comparison between a paper and a paper identified as a potential match and the generation of an originality report based upon the full text comparison

analysis to report on the degree of similarity. The passage cited by the Examiner in the Broder patent, however, only describes detection of similarity (and not the degree of), based upon a comparison of representative selections of documents, and no subsequent analysis of the full documents. As such, the Examiner's support for the motivation to combine the Carden and Broder patents fails.

The Examiner further suggests that "Broder teaches a process of checking for resemblance of documents in a Plagiarism and Intellectual property situation where authors can detect copies of their own work using a resemblance technique that employs a containing/containment technique that can possess a full document to full document comparison." The Applicants disagree with this assessment of Broder. A resemblance technique is the comparison, as stated above and found throughout Broder, of representative selections of a document. A containing/containment technique or metric is well known in the art and based on the same principle of comparison of representative selections that can be further calculated by an algorithm for estimating the probability of the representative selection (e.g. fingerprint) of document A being roughly contained in the chosen selection (e.g., fingerprint) of document B (column 8, lines 6-7). Broder does not state that full text papers are compared using a contained/containing technique, as Broder only teaches the comparison of representative selections of documents. As well, Broder does not suggest that an originality report of results of the chosen technique is generated to present to a user the probability of resemblance or containment.

Considering that the cited references do not teach or suggest a system comprising an originality checking application configured to conduct a full text comparison between papers and generate an originality report based on the full text comparison, one skilled in the art would not be motivated to combine the Broder patent with the Carden patent in the manner suggested by the Examiner.

The Examiner lastly states that Carden and Broder are "analogous art because they are from the same problem solving area of processing documents on the internet and relating documents to a plurality of documents." (Final Office Action, page 18). The Applicants disagree, and suggest that Broder and Carden are non-analogous art, with the only tie between the two cited references being that they are computer-based methods. Carden teaches a computer based publishing system to accept, manage, review (manually

by reviewers), track and publish manuscripts and other forms of media that are submitted, such that a person can submit a manuscript, for example, for consideration for publication in a journal, conference proceeding, etc. The processing of documents by Carden is not done on the internet, and no relating of documents to a plurality of documents for plagiarism purposes is performed or even contemplated. The submission and communication between the subscription parties is performed over the internet, however the actual process is done on the server that is either hosted on a service provider's computer that allows access to all involved parties to that site, or the system is installed on a dedicated server system at a particular publishing house. Broder, on the other hand, teaches the indexing of Web sites performed by a search engine on the internet and solely for the purpose of indexing similar Web pages, thereby providing a user a list of sites that, by application of an algorithm based on rough comparisons, are found to match with a certain probability the terms as entered by the person performing the query. This is akin to comparing apples and oranges; they are both fruit, but they are not analogous.

As such, the Applicants request these rejections be withdrawn.


II. Claims 53 and 58 are not obvious under 35 U.S.C. §103(a) over Carden in view of Evans and further in view of Gehringer

The Examiner has rejected dependent claims 53 and 58 under 35 U.S.C. §103(a) as allegedly being unpatentable over Carden in view of Evans, and further in view of Gehringer. Claims 53 and 58 are dependent upon non-obvious and non-anticipated claim 43. As discussed in Section I of this Final Office Action Response, the Carden patent fails to teach all of the required elements within claim 43. The other cited references do not remedy this deficiency. As such, a *prima facie* case of obviousness has not been established. The Applicants request the rejection be withdrawn.

CONCLUSION

If a telephone interview would aid in the prosecution of this application, the Examiner is encouraged to call the undersigned collect at (608) 218-6900.

Dated: 11/2/2006



Lisa G. de Berg
Registration No. 56,524

MEDLEN & CARROLL, LLP
101 Howard Street, Suite 350
San Francisco, California 94105
(608) 218-6900